BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Software

Volume 8(14)

RDNAnalyzer: A tool for DNA secondary structure prediction and sequence analysis

Muhammad Afzal[1, 2], Ahmad Ali Shahid[1], Abida Shehzadi[1]*, Shahid Nadeem[2] & Tayyab Husnain[1]

[1]Centre of Excellence in Molecular Biology, University of the Punjab, Lahore, Pakistan; [2]Department of Bioinformatics and Biotechnology, GC University, Faisalabad, Pakistan; Abida Shehzadi – Email: abida_ravian@hotmail.com; *Corresponding author

Received June 26, 2012; Accepted July 05, 2012; Published July 21, 2012

Abstract:
RDNAnalyzer is an innovative computer based tool designed for DNA secondary structure prediction and sequence analysis. It can randomly generate the DNA sequence or user can upload the sequences of their own interest in RAW format. It uses and extends the Nussinov dynamic programming algorithm and has various application for the sequence analysis. It predicts the DNA secondary structure and base pairings. It also provides the tools for routinely performed sequence analysis by the biological scientists such as DNA replication, reverse compliment generation, transcription, translation, sequence specific information as total number of nucleotide bases, ATGC base contents along with their respective percentages and sequence cleaner. RDNAnalyzer is a unique tool developed in Microsoft Visual Studio 2008 using Microsoft Visual C# and Windows Presentation Foundation and provides user friendly environment for sequence analysis. It is freely available.

Availability: http://www.cemb.edu.pk/sw.html

Abbreviations: RDNAnalyzer (Random DNA Analyser), GUI: Graphical user interface, XAML (Extensible Application Markup Language)

Keywords: Sequence analysis, DNA, nucleotides, Nussinov algorithm, C# language

Background:
One of the most significant biological discoveries of 20th century was the discovery of structure and function of DNA molecule that posed the tremendous impact on science and medicine **[1]**. Identification of disease causing genes **[2]** and the pharmaceutics to cure the disease tremendously changes the visage of future science. Science of modern medicine and genetics are thoroughly based on the DNAs to diagnosis disease, future genetic predisposition of disease, gene therapy and new drug discovery are all based on individual genetic profiles. It thus presents hope for past incurable genetic diseases **[3]**. Additionally gene cloning for disease resistance plants and animals **[4]**, to solve the various cases of forensic science as one's true paternity, victim and crime suspects **[5]** and last but not the least DNA microarray based studies have opened a new era of personal genomics etc. are being largely solved because of the understanding and knowledge of DNA structure and function.

In short, DNA molecules are used for purposes that go beyond their functions in nature. The mother of all DNA based studies is Polymerase Chain Reaction (PCR) **[6].** PCR is the involved at any stage of the applications of DNA based studies. PCR requires primers in any of its application starting from DNA marker based studies to sequencing and Real Time PCR. Primers are single stranded DNA molecules generated in sequence specific manner to amplify only the DNA of interest. Single strand DNA either in the form of primer or oligos

(microarray probes) molecules fold into specific three-dimensional conformations determined by nucleotides sequence, thereby lowering the so-called "free energy" of the molecule. The lower is the free energy; the stable will be the secondary structure. Various interactions pose their role for DNA confirmations forming hairpins, pseudoknots and triple helices. Prediction of single stranded DNA structure can be used for optimized primer design to avoid self folding. Such predictions are also important for identification of single strand conformational polymorphism, DNA tags and error prone chemical and enzymatic interactions **[7]**.

Since bio-molecular function follows from its shape and structure, knowing that shape is invaluable in endeavors such as creating new drugs and understanding genetic diseases. Our current physical methods (X-Ray Crystallography and Nuclear Magnetic Resonance) are too expensive and time consuming, so a hot topic in bioinformatics is structure prediction. The idea is we take the sequences of bases which make up a biomolecule such as single strand DNA words, and try to determine how that sequence folds to form the final shape or structure.

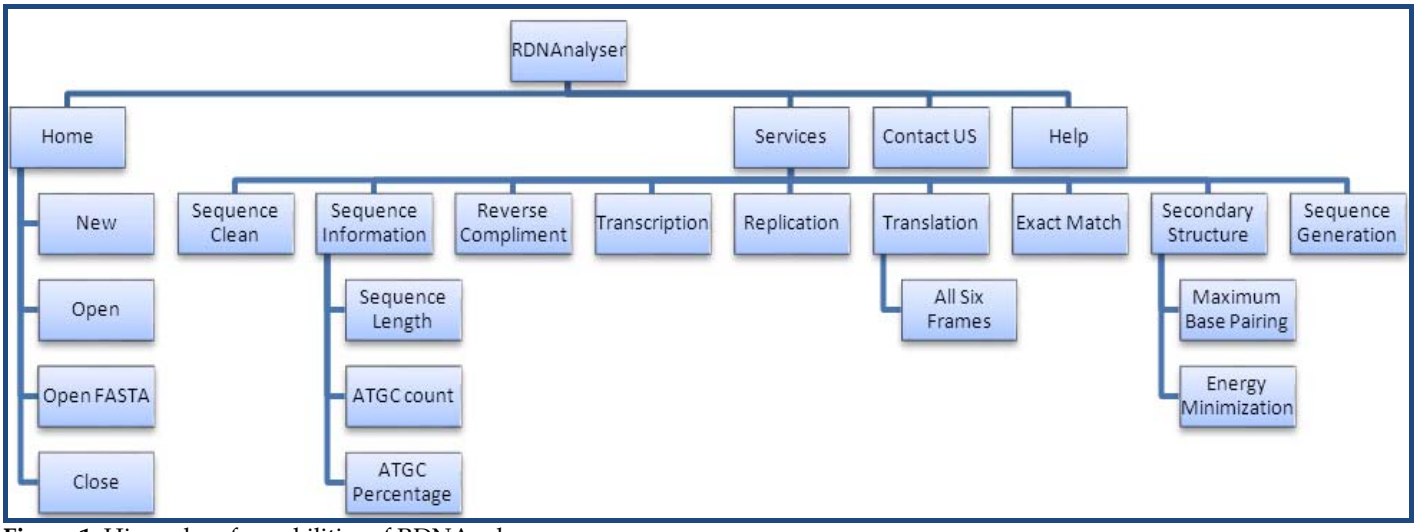

Figure 1: Hierarchy of capabilities of RDNAnalyzer

Implementation of RDNAnalyzer:

This RDNAnalyzer was designed after complete analysis of existing DNA analysis tools. The major drawback of those softwares is the web based application. That's why their access is only on the availability of internet. It was designed as desktop application to overcome this drawback. Moreover, various routinely used tools (e.g. DNA secondary structure prediction and primer looping **[8]**, replication, transcription, translation **[9]**, GC content information in the given nucleotide sequence **[10]**, etc.) were collected at single platform with ease of use and requiring not any computer skills of user's or knowledge seekers. It uses the Nussinov dynamic programming algorithm **[11]**.

Different tools were used for the software development like Microsoft Visual Studio 2008 (Professional Edition) and Microsoft .Net and Visual C#. Microsoft Visual Studio 2008 has self-generated code that in RDNAnalyzer and it saved time. The .NET Framework provides applications for programmers to work with languages, devices and services of their own choice **[15].** In this research project .Net 3.5 is used which is by-default available with Microsoft Visual studio 2008. C# is a programming language that is easy to use and have good qualities. Many bioinformatics tools are developed in this language. It can create windows applications by using Microsoft Visual studio. The graphical user interface (GUI) of RDNAnalyzer is designed in WPF (Windows Presentation Foundation). It represents the basic features of a GUI application, including main windows, dialog boxes, controls, menu systems, and others. It is simple but powerful object model.

RDNAnalyzer-Interface:

RDNAnalyzer was basically developed for the secondary structure analysis of the DNA generated from word generating techniques. It finds suitable complimentary base pairs of the available sequence. RDNAnalyzer interface was designed by the XAML (Extensible Application Markup Language) coding. The Home page of RDNAnalyzer has options like New (to start a new work), Open (to open a sequence containing text file), Open Fasta (to open Fasta format files) and Close (to close all applications). DNA secondary structure prediction, replication, transcription, translation, exact match, etc. are available in services page. User can provide the DNA sequence of their interest at new work at home page or can open the sequence file via browse button or a FASTA file. Clean sequence button cleans the DNA sequence info and provides ATGC's for further analysis. DNA secondary structure can be predicted at maximum base pairing or at minimized energies. The predicted report can be saved as such in a text file or it can be further extended to draw the structure where the respective bases form complimentary structures and then respectively be saves as image file for future use. Similarly other functions performed with the RDNAnalyzer are shown in **(Figure 1)**.

Discussion:

The objective of developing the RDNAnalyzer is to provide a system based tool to allow researchers and users to quickly perform various routinely used applications of DNA sequences. The basic purpose of this software development is to predict the secondary structure of DNA single strand. It can also successfully replicate the DNA sequence, reverse compliment generation, transcription, and translation. This software also

detects the maximum base pairs in a DNA sequence. It provides a data structure that supports a fast and efficient analysis method. The algorithms used in this software could also be implemented in other areas of work.

The important application of RDNAnalyzer are prediction of secondary structure by using sequence in FASTA format; analysis of replication, transcription and translation of a DNA sequence; analysis of GC content ratio and percentages of nitrogenous base and to draw positions of base pairs. It works in Microsoft Windows version and hence, provides user friendly environment for sequence analysis. Some of the important services provided by RDNAnalyzer are:

DNA secondary structure prediction

The secondary structure prediction is shown in dot format in which dots (.) represent the bases which do not have base-pairing and braces represent the nucleotides that form base pair with each other **(see supplementary material)**.

DNA to protein

Conversion of nucleotide sequence to protein sequence is routinely used application by the biological sciences researchers. RDNAnalyzer can convert the DNA into RNA and 6 reading frames of proteins i.e. 3 forward frames and 3 reverse frames.

Sequence Composition/sequence info

Another important application of RDNAnalyzer is to provide the detailed information about the nucleotide sequence of DNA i.e. the total number of nucleotides and the number of bases and their respective percentages. Information about the percentages of bases is very important especially the GC contents for optimized primer design.

The graphical user interface of RDNAlyzer is very simple, straight forward and easy to follow, fast and gives efficient output. It checks either user entered a valid sequence or not. It can reduce time and effort as compared to other tools during different research projects. Currently it can be downloaded in format for window based applications. Our future implementation of the software is to make it online with more accuracy and functionality including the use of DNA sequences in different formats, online search of the required sequences and their downloading and respective uses in any of the available format. Another important goal of the software is to make it compatible for Linux environment to provide the benefit to all types of researchers who prefers to work on window, Linux or web pages.

Requirements:

Project name: RDNAnalyzer; Operating system(s): Windows based PC; Programming language: C # and Windows Presentation Foundation.

Competing interests:

The authors declare that they have no competing interests.

Authors contributions:

MA and AS designed the study. MA designed the tools and SN helped in script coding. AS drafted the manuscript. AAS supervised the research. AAS and TH critically reviewed the manuscript. All authors have read and approved the final manuscript.

References:

[1] McCarty M, *Nature.* 2003 **83**: 89 [PMID: 12540908]
[2] Debouck C & Goodfellow PN, *Nat Genet.* 1999 **21**: 50 [PMID: 9915501]
[3] Kumar D,*Ann Ist Super Sanita.* 2011. **47**: 31[PMID: 21430336]
[4] McDowell JM *et al. Trends Biotechnol.* 2003 **21**: 178 [PMID: 12679066]
[5] McDowell JM & Woffenden BJ, *Stat Sci.* 1991 **6**: 175
[6] Schochetman G *et al. J Infect Dis.* 1988 **158**: 1154
[7] Dong F *et al. Nucleic Acids Res.* 2001 **29**: 3248 [PMID: 11470883]
[8] Horisaka T *et al. J Clin Microbiol.* 2004 **42**: 5349 [PMID: 15528740]
[9] Kumaria R *et al. Virol J.* 2011 **8**: 372 [PMID:21794174]
[10] Shehzadi A *et al, Virol J.* 2011. **8**: 55 [PMID:21303499]
[11] Nussinov R & Jacobson AB, *Proc Natl Acad Sci.* 1980 **77**: 6309 [PMID: 6161375]

Edited by P Kangueane
Citation: Afzal *et al.* Bioinformation 8(14): 687-690 (2012)

BIOINFORMATION

open access

Supplementary material:

DNA secondary structure prediction:
The secondary structure prediction is shown in dot format in which dots (.) represent the bases which do not have base-pairing and braces represent the nucleotides that form base pair with each other.
The report of secondary structure prediction is given as:

Job Title: C:\Documents and Settings\Administrator\Desktop\seq1.fasta
Sequence Detail: (If Available)

Available sequence relevant information:
>gi|262205176|ref|NR_030166.1| Homo sapiens microRNA 491 (MIR491), microRNA
By using Nussinov algorithm Maximum Base Pairing implementation

Given DNA Sequence: (Sequence Length: 84).
TTGACTTAGCTGGGTAGTGGGGAACCCTTCCATGAGGAGTAGAACACTCCTTATGCAAGATTCCCTTCTACCTGGCTGGGT
TGG
Secondary structure:
...((..(((..((((((..((((((.(.((.(((.((((((.....)))))).))).))).))))))..))))))..)))..))...
****